Counterfactual Consent and the Use of Deception in Research

bioethics

Bioethics ISSN 0269-9702 (print); 1467-8519 (online) doi:10.1111/bioe.12142

COUNTERFACTUAL CONSENT AND THE USE OF DECEPTION IN RESEARCH

ALAN T. WILSON

Keywords
deception,
consent,
research ethics

ABSTRACT
The use of deception for the purposes of research is a widespread practice within many areas of study. If we want to avoid either absolute acceptance or absolute rejection of this practice then we require some method of distinguishing between those uses of deception which are morally acceptable and those which are not. In this article I discuss the concept of counterfactual consent, and propose a related distinction between counterfactual-defeating deception and counterfactual-compatible deception. The aim is to show that this proposed distinction will be useful in furthering the debate regarding the use of deception for the purposes of research.

INTRODUCTION

The use of deception for the purposes of research has long been the subject of interest and concern from ethicists and researchers alike. Many studies require that participants be deceived about various factors, and this raises interesting questions about the moral acceptability of those studies. If we want to avoid either absolute rejection or absolute acceptance of this practice, then we face the challenge of providing some method for distinguishing between different cases. In response to this challenge, some have argued that it is important to distinguish between different *types* of deception, and that making this distinction means that we can then judge the moral acceptability of a given study depending on the type of deception used. In this article I will highlight a distinction between different forms of deception that appeals to the notion of *counterfactual consent*. I aim to show how this way of distinguishing between types of deception is superior to other suggestions in the recent literature, and to demonstrate how it can shed light on the issue of when the use of deception for the purposes of research ought to be considered morally acceptable.

1. THE USE OF DECEPTION IN RESEARCH

Deception can be said to have taken place whenever an agent communicates in such a way as can be reasonably expected to result in a false belief.[1] And the use of deception in various fields of research is fairly commonplace, as has been noted throughout the literature on this topic.[2] For the purposes of this article, it will be useful to consider particular examples and to highlight some of the different ways in which deception can occur. Some examples of studies involving deception have recently been provided by McCambridge et al.[3] In one of these studies,

[1] D. Wendler & F.G. Miller. Deception in Clinical Research. In: E.J. Emanuel et al., editors. *The Oxford Textbook of Clinical Research Ethics*. New York, NY: Oxford University Press; 2008. p. 315–324, 316.
[2] See, for example, J.H. Korn. 1997. *Illusions of Reality: A History of Deception in Social Psychology*. Albany, NY: SUNY Press; and A. Ortmann & R. Hertwig. Is Deception Acceptable? *Am Psychol* 1997; 52: 746–747.
[3] J. McCambridge, K. Kypri, P. Bendtsen & J. Porter. The Use of Deception in Public Health Behavioral Intervention Trials: A Case Study of Three Online Alcohol Trials. *AJOB* 2013; 13: 39–47.

Address for correspondence: Alan T. Wilson, University of Edinburgh, Dugald Stewart Building, 3 Charles Street, Edinburgh EH8 9AD. Email: a.t.wilson@sms.ed.ac.uk
Conflict of interest statement: No conflicts declared

participants believed themselves to be involved in a review of general health and well-being when, in fact, researchers were exclusively interested in their habits of alcohol consumption.[4] That is, the subjects were deceived about the true purpose of the research in which they were participating, or *why* the experiment was being carried out. A second example of the use of deception in research can be taken from the work that has fuelled the recent 'situationist' movement in psychology and philosophy. In many of these experiments, situations are manufactured in order to observe how members of the public will react, and to determine whether or not subtle differences in how the situation is presented will alter those reactions.[5] In one such experiment, a member of the research team gave the impression of having accidently dropped some papers in front of a stranger while walking through a shopping centre. Researchers then noted how often members of the public would stop to help collect the papers, and how the frequency of helpful behaviour varied with seemingly inconsequential features of the environment.[6] In this example, then, steps were taken to mislead the subjects about *whether* the research was even taking place, as well as about certain other *matters of fact*, such as that the papers had been dropped accidentally. For our final example we can use perhaps the most famous case of deception in research – the Milgram experiments. In these experiments, subjects were asked to administer electrical shocks to another participant for the purposes of research into the effects of punishment on learning behaviours.[7] However, the true purpose of the experiments was to investigate obedience tendencies. The shocks were not real and the other participant was in fact an accomplice of the researchers. Therefore, in addition to being deceived about why the experiment was taking place, subjects were also deceived about certain other matters of fact (such as whether or not the shocks were legitimate or whether or not the victim was a fellow volunteer). These examples highlight some of the ways in which deception can be carried out – focusing on either the purpose of the research, certain features of the research environment, or even on whether or not an experiment is taking place. An important question is how we ought to react to this widespread practice – how does the use of deception impact on the moral acceptability of a given study?

Moral absolutism regarding two separate questions ought to be ruled out at this stage. The first question regards how we ought to view the impact of problematic deception on the moral acceptability of a study. An absolutist might think either that any study involving problematic deception is automatically unjustifiable or that any study lacking problematic deception is automatically acceptable. Both of these positions should probably be rejected, although the latter is more obviously wrong. There are many factors that can render a study morally unacceptable, and the use of problematic deception is only one of these. A given study may avoid problematic deception and yet still be unjustifiable on other grounds, such as the exposure of participants to unacceptable risks, the violation of appropriate consent, or perhaps even the pursuit of intrinsically immoral research goals. The contrary absolutist response to this question – that any use of problematic deception renders a study unjustifiable – is more difficult to dismiss. This is partly because we do not yet have an understanding of what is meant by 'problematic' deception. How this phrase is to be interpreted will affect the acceptability of this form of absolutism. However, it seems likely that, on most interpretations, the absolutist line ought to be avoided. It is at least theoretically possible that the benefits of a given study could be sufficiently high, and the level of problematic deception sufficiently low, such that the study would be justifiable overall. Perhaps any such example would be unrealistic, but the possibility should prompt us to be cautious about what we can expect from the notion of 'problematic' deception. It is not necessarily the case that all studies featuring problematic deception will be unjustifiable overall, or that all studies lacking such deception will be acceptable overall. Considerations stemming from deception alone will not always be sufficient to settle the matter, and to expect this from a conception of problematic deception would be to set the bar too high. We should instead expect a useful conception of problematic deception to identify those cases where the use of deception is *prima facie* problematic. It ought to provide a method for identifying those studies where there is a special burden of justification (stemming from considerations of deception) on those who want the study to be permitted. My proposed distinction will aim to provide such a useful conception.

The second question to which absolutism would be an inappropriate response concerns which forms of deception ought to be considered problematic. The view that all deception is problematic deception is not plausible. For example, in certain placebo trials efforts will be taken to make participants believe that they might be receiving the actual drug, even after it has been determined that a placebo will be administered. To that extent there is a form of deception taking place, but this deception does

[4] Ibid: 41–43.

[5] See J. Doris. *Lack of Character*. Cambridge: Cambridge University Press; 2002; L. Ross & R. Nisbett. *The Person and the Situation*. Boston, MA: McGraw Hill; 1991; D. Russell. *Practical Intelligence and the Virtues*. Oxford: Oxford University Press; 2009. Ch 9.

[6] A.M. Isen & P.F. Levin. Effect of Feeling Good on Helping: Cookies and Kindness. *J Pers Soc Psychol*. 1972; 21: 384–388. This case is detailed in R. Adams. *A Theory of Virtue*. Oxford: Oxford University Press; 2006. p. 117–118.

[7] S. Milgram. Behavioral Study of Obedience. *J Abnorm Soc Psychol* 1963; 67: 371–378.

not appear to be problematic.[8] We will consider later exactly why cases of this sort seem unproblematic but, whatever the explanation, the possibility of such cases tells against the absolutist view that all deception be considered problematic. And we should also reject the contrary absolutist response to this question whereby deception is never problematic. Imagine a study where participants are informed that researchers will be observing their behaviour but are deceived into believing that this process will end once they leave the laboratory. In fact, the researchers follow the participants home and spy on them through the window. This would appear to be a morally problematic use of deception, and any researcher wishing to carry out such an experiment might find themselves hard-pressed to justify its use.[9] Providing real-life cases of obviously problematic deception is more controversial, although many have argued that the aforementioned Milgram experiments are of this sort.[10] Either way, the very possibility of problematic deception is unlikely to be questioned. Both absolutist responses to this question ought, therefore, to be avoided. And if some forms of deception are problematic while others are not, it will be useful to have some method for working out how particular cases ought to be classified. I will now consider some of the attempts in the recent literature to separate those types of deception that are *prima facie* acceptable from those that are *prima facie* unacceptable. I will argue that these attempts are less helpful than might have been hoped.

2. DISTINGUISHING BETWEEN DIFFERENT TYPES OF DECEPTION

One way of responding to this issue is by appeal to the notion of autonomy. We can distinguish forms of deception depending on whether or not they undermine the autonomy of the participants. Of course, as Nomy Arpaly has highlighted, the term 'autonomy' has been used in many different ways by many different theorists.[11] When assessing this approach it is important to be clear about which understanding of autonomy we have in mind. Rachel Zuraw appeals to a Kantian understanding and then applies this to the problem at hand.[12] On the Kantian understanding, respect for autonomy requires that any rational agent ought always to be 'treated as an end in and of his or her self, and not merely as a means towards fulfilling others' goals.'[13] Zuraw aims to provide a method of distinguishing types of deception that highlights the greater tendency of one form to infringe upon an agent's autonomy. The suggestion is that we distinguish between cases where the subject is deceived about the *goals* of research and cases where the subject is deceived about *whether or not* the study is even taking place. As we saw above, the Milgram experiments and the alcohol study described by McCambridge et al. would be examples of the former. Cases from the situationist literature would provide examples of the latter. Zuraw argues that deception about goals does not violate autonomy but that deception about the existence of the study does, and that this will be a useful distinction for determining the acceptability of a study that uses deception. Zuraw states that:

> Provided that the research participant has not been misled about something that would actually cause harm to her, a Kantian could argue that she had all of the information necessary . . . Conversely, a person who is studied without her knowledge has no opportunity at all to consider whether her goals might be served by participating in research, and is being used merely as a tool for the researchers' ends.[14]

However, the distinction suggested by Zuraw will not work. It is possible for subjects to be treated as mere means even when the deception is (only) about the goals of the research. Imagine a case where an agent would be willing to participate in a study seeking to further the understanding of certain cancer treatments, but would not be willing to participate in a study focused on the treatment of any other illness or disorder. With this information in hand, it would be possible for researchers to deceive the agent into believing that their study aims to further the understanding of cancer treatments, and to secure the agent's participation as a result of this deception. In such a case, it is clear that the agent's autonomy is not being respected. It is also intuitively plausible that this use of deception would be morally problematic. Therefore, while some understanding of autonomy may yet prove useful for our purposes, the distinction suggested by Zuraw will not be sufficient. The attempt to uncover problematic uses of deception by distinguishing

[8] The issue of whether deception is involved in placebo trials (and to what extent) is covered in Wendler, Miller, *op. cit*. note 1, p. 316.

[9] One (surprisingly) real example of research where subjects where spied on without their knowledge is provided by Korn, *op. cit*. note 2, p. 3. In this case subjects were spied on while using a public urinal 'by a person using a periscope to peek under a toilet stall'.

[10] See: D. Baumrind. Some Thoughts on Ethics of Research: After Reading Milgram's 'Behavioral Study of Obedience'. *Am Psychol* 1964; 19: 421–423; H.C. Kelman. Deception in Social Research. *Trans Action* 1966; 3: 20–24; also the references in D. Mixon. Instead of Deception. *J Theory Soc Behav* 1972; 2: 145–177. For a recent review of these experiments and of their ethical status see G. Perry. *Behind the Shock Machine: The Untold Story of the Notorious Milgram Psychology Experiments*. New York: The New Press; 2013.

[11] N. Arpaly. *Unprincipled Virtue*. Oxford: Oxford University Press; 2002. Ch. 4.

[12] R. Zuraw. Consenting in the Dark: Choose Your Own Deception. *AJOB* 2013; 13: 57–59.

[13] Ibid: 58.

[14] Ibid.

between deception about why the research is taking place and deception about whether research is taking place is not as helpful as might have been hoped.[15]

Another theorist who highlights the importance of autonomy in this regard is C.D. Herrera.[16] In 'Ethics, Deception and "Those Milgram Experiments" ', Herrera argues that deception does not impact on an agent's autonomy, and is therefore less problematic, when the agent has consented to being deceived in advance. Herrera states that:

> we can agree that when psychologists deceive subjects with no advance warning they are using those subjects as means to an end that the subjects cannot share . . . In medical research, subjects in a clinical trial might participate knowing that researchers are purposely withholding some information from them. In the same way, psychologists could tell subjects in a psychology experiment that they are deliberately withholding or even misrepresenting details about the study. If a subject understands that she is consenting on the basis of incomplete information, her autonomy isn't violated.[17]

This suggestion – that *prima facie* unproblematic deception will be deception with prior consent – has since been repeated elsewhere.[18] And the suggestion brings with it certain benefits. First of all, it allows us to explain the intuition that the deception used in certain placebo trials (as mentioned above) is not problematic. Participants will have been made aware of the possibility that they will receive a placebo and that members of the research team may perform actions designed to conceal from them the truth of whether or not they are receiving the actual drug. Given that the participant consents to this in advance, any subsequent deception appears less problematic than it otherwise would have. A further benefit of Herrera's suggestion is that it leaves open the possibility that a study where advanced consent for deception has been received may nevertheless be morally unacceptable on other grounds. It was suggested above that a realistic conception of problematic deception would not entirely answer the question of whether a given study is acceptable or not. A study may yet be unacceptable on the grounds that the levels of risk are too high or that the consent offered was not sufficient, and the suggestion from Herrera is compatible with this fact. For these reasons, the suggestion that we use the notion of advanced consent to determine whether a participant's autonomy is violated and, therefore, whether the deception used is problematic, appears to be a step in the right direction.

However, this way of responding to the issue cannot do all of the work that is required. If we focus solely on whether or not the participant has actually consented in advance to being deceived then we will be left with certain unanswered questions. This is because there will be some studies where receiving advanced consent to be deceived will be incompatible with the study itself. In some cases it will not be possible to ask participants to consent in advance to be deceived due to concerns that this will render them suspicious and so less likely to behave naturally during any experiment. In other cases it will not be possible to ask participants to consent in advance *at all*, such as in the situationist studies mentioned above where members of the public are observed reacting to manufactured scenarios. The current suggestion makes it the case that it would be impossible to carry out any of these studies without engaging in problematic deception, and this result will be unintuitive to many. Furthermore, even if we were inclined to accept this verdict, it would still be useful if our method for distinguishing forms of deception would allow us to differentiate between examples of this sort. Therefore, despite the benefits of accepting Herrera's suggestion, we have reason to look for some other distinction. A distinction appealing to a need for *actual* consent will not be sufficient for our purposes.

3. COUNTERFACTUAL CONSENT

An agent can be said to have given their counterfactually robust consent, or *counterfactual consent*, when the following two conditions are satisfied: (1) they consent, and (2) they are not ignorant of any fact that would have caused them to withhold their consent if they were aware of it. The notion of counterfactual consent can be made clearer with an example. Suppose that a colleague asks for permission to put a photograph of me 'on the website'. Assuming that they are referring to the departmental website, and assuming that they are acting in a professional capacity, I consent. However, unbeknownst to me, they were actually referring to a website entirely unrelated to our professional lives. In this case, the first condition for counterfactual consent has been met – I offered my consent – but it is not yet clear whether the second condition has been satisfied. This will depend upon how awareness of the true nature of the website would have impacted on my willingness to consent. If this knowledge would have led me to refuse to consent then

[15] A further worry with Zuraw's suggestion is that it has nothing to say regarding cases of deception which are about neither the goals of the study nor the existence of the study. As we saw above, some studies deceive participants about other matters of fact in order to observe their reactions.

[16] C.D. Herrera. Ethics, Deception, and 'Those Milgram Experiments'. *J Appl Philos* 2001; 18: 245–256, p. 250.

[17] Ibid.

[18] L. Bortolotti & M. Mameli. Deception in Psychology: Moral Costs and Benefits of Unsought Self-Knowledge. *Account Res*. 2006; 13: 259–275, p. 266; D. Wendler & F.G. Miller. Deception in the Pursuit of Science. *Arch Intern Med* 2004; 164: 597–600, p. 598; Wendler, Miller, *op. cit*. note 1, p. 320–321.

the second necessary condition has not been met and we should say that my colleague did not have my counterfactual consent. On the other hand, if I would have been just as happy to consent upon uncovering the true nature of the website then we can say that counterfactual consent had in fact been secured. This example therefore helps to explain the two necessary conditions for counterfactual consent.

It is important to highlight the possible relationship between counterfactual consent and the more widespread notion of informed consent. It is not the aim of this article to provide a set of necessary and sufficient conditions for informed consent, a notion that has been understood in many different ways throughout the literature.[19] However, it is important to point out that some theorists have included what I have listed as the second necessary condition for counterfactual consent within their conceptions of informed consent. For example, Wendler and Miller argue that an agent should not be considered to have offered informed consent if they are unaware of 'any aspects of the study that would affect their willingness to participate'.[20] It may be, therefore, that what I am calling counterfactual consent is in fact one of the necessary components of informed consent. However, it ought to be acknowledged that counterfactual consent will be a less demanding notion than informed consent, as is shown by the example above. Regardless of how I would have responded upon learning of the true nature of the website, it is clear that, at the time of offering my consent, I was not sufficiently informed. I was ignorant of an important and relevant detail and so could not offer my informed consent. However, we have seen that it was possible for me to offer counterfactual consent, so long as I *would* have consented upon becoming informed. Therefore, the conditions for counterfactual consent are less demanding than those for informed consent, and this article aims to suggest that progress on the issue of using deception in research can be made by appealing to the less demanding notion. Similarly, Wendler and Miller include the demand that 'people are not deceived about aspects of the study that would affect their willingness to participate, including risks and potential benefits' as just one component in their list of restrictions on the use of deception.[21] Other necessary conditions in order for deception to be acceptable include the requirement that participants consent to the deception in advance (in a way similar to Herrera's suggestion) *and* that the full extent of the deception be revealed to participants following the study.[22] Again, the aim of this article is to suggest that appealing to a less demanding set of restrictions will be sufficient in order to advance the debate regarding whether or not a given instance of deception is itself problematic.

One further important aspect worth clarifying is that counterfactual consent does require actual consent. The first necessary condition for counterfactual consent is that the agent actually consents. This means that we cannot correctly say that we have an agent's counterfactual consent if they have not actually consented – regardless of whether they are fully informed or whether they would have consented if asked. We saw above that any appeal to *actual* consent will be insufficient for the purposes of identifying problematic forms of deception. This is because it would automatically rule out studies where securing actual consent is not possible – such as in the situationist experiments – and would also leave us with no resources with which to distinguish between cases of this sort. Therefore, because counterfactual consent is a form of actual consent, we should not identify problematic deception as deception lacking counterfactual consent. Instead, my proposal is that a useful method for distinguishing forms of deception can be provided by focusing on just the second necessary condition for counterfactual consent.

We can distinguish types of deception depending on whether or not awareness of the information being concealed would have caused the subject to refuse their consent (regardless of whether or not that consent has actually been received). If this is the case, then we can say that we have an example of *counterfactual-defeating deception*. On the other hand, if awareness of the concealed information would not cause an agent to refuse their consent, then we can say that we have an example of *counterfactual-compatible deception*. In the example above, imagine that my colleague purposely led me to form my incorrect belief about the nature of the website that would host my photograph. If awareness of the facts would have led to my refusing to consent, then my colleague is guilty of counterfactual-defeating deception. If this is not the case then my colleague has carried out counterfactual-compatible deception. And this fact about the nature of the deception will be true regardless of whether or not consent has actually been offered.[23] My suggestion is that the distinction between

[19] E. Blacksher & J.D. Moreno. A History of Informed Consent in Clinical Research. In: E.J. Emanuel et al., editors. *The Oxford Textbook of Clinical Research Ethics*. New York: Oxford University Press; 2008. p. 591–605.

[20] Wendler, Miller, *op. cit*. note 1, p. 321. See also Wendler, Miller, *op. cit*. note 18, p. 599. The importance of counterfactual robustness for consent has also been highlighted recently in connection with a separate issue within applied ethics (although not using this terminology). See T. Dougherty. Sex, Lies, and Consent. *Ethics* 2013; 123: 717–744.

[21] Wendler, Miller, *op. cit*. note 1, p. 323

[22] Ibid.

[23] The distinction here is different from that offered by Collin O'Neil between 'methodological manipulation' and 'inducement manipulation'. The definition of those terms focuses on the intentions or 'purposes' of the deceiver, whereas the proposal here makes no such reference. The distinction proposed here would therefore generate different verdicts from O'Neil's proposal. See C. O'Neil. Methodological and Inducement Manipulation. *AJOB* 2013; 13: 55–57.

counterfactual-defeating and counterfactual-compatible deception can allow us to identify *prima facie* problematic forms of deception. In order to show that this is the case, we can now reconsider some of the examples set-out above.

4. APPLYING THE PROPOSED DISTINCTION

If we think that deception can sometimes be an acceptable practice for the purposes of research then we have good reason to consider whether or not the moral acceptability of a given study depends on the type of deception that is carried out. I believe that the notion of counterfactual consent leads us towards a useful way of approaching this issue. Researchers (and ethicists evaluating research methods) ought to distinguish between counterfactual-defeating deception and counterfactual-compatible deception. Cases of counterfactual-defeating deception ought to be viewed with greater suspicion than cases of counterfactual-compatible deception, and will therefore generally require a higher level of justification if they are to be permitted. It will be useful to reconsider the examples of deception that were set-out above in order to show how this suggestion is in-line with intuitively acceptable verdicts in specific cases.

In order to avoid controversy, I have provided two fictional examples in order to demonstrate some *prima facie* unacceptable uses of deception in research. First, there was the example where participants were deceived regarding the length of time for which they would be observed, with researchers then following them home and spying on them through the window. This looks to be an unacceptable use of deception, and the suggested method agrees with that verdict. Awareness of the fact that they will be followed home and spied on appears likely to render a participant unwilling to give their consent. We therefore have an example of counterfactual-defeating deception and are rightly suspicious about the moral acceptability of such a case. Secondly, in response to Zuraw's suggested distinction, I provided an example where a participant was deceived about the purposes of a study. Knowing that the agent would only consent to participate in research with a particular focus, researchers deceived the agent into believing that their study shared that focus. This example seems particularly problematic, and the proposed distinction can tell us why. We have here a clear case of researchers (knowingly) engaging in counterfactual-defeating deception. And we would therefore be justified in asking serious questions about the moral acceptability of their study. What these examples show is that the proposed distinction between counterfactual-defeating and counterfactual-compatible deception provides us with the intuitively correct results in cases where the use of deception appears unacceptable.

The proposed method also gets things right in the most obvious cases of morally acceptable deception. Consider again the placebo trials where participants know in advance that they may be deceived into believing that they are receiving the actual drug. We saw above that Herrera's appeal to advanced consent is one way of explaining the acceptability of this example. Participants consent in advance to the subsequent deception and this explains why that deception is unproblematic. However, the current proposal is also sufficient for explaining this fact. By *actually* consenting in advance to the possibility that they will be participating in a study while receiving a placebo, the participant provides strong evidence that they *would* consent to participate were they to be aware of the fact that they were receiving a placebo. We therefore have good reason to believe not only that the deception will be counterfactual-compatible, but that counterfactual consent has actually been offered. Both of the necessary requirements for counterfactual consent appear to be satisfied. (Notice that this does not show that informed consent has been offered, as there may yet be some other relevant information of which the participant is not aware.) By consenting in advance to the subsequent deception, the participant shows that the deception will be counterfactual-compatible. The proposed distinction is therefore able to accord with intuitions regarding cases of obviously morally permissible deception, as well as cases of obviously morally problematic deception. We should now consider cases where the acceptability of deception is less obvious.

The three examples listed at the start of this paper are all cases where the moral acceptability of a study is open to question (for those who believe that deception is at least sometimes permissible). I have already mentioned that the use of deception in the Milgram experiments has been criticized by some while defended by others. The deception present in the experiments relied on by the situationist movement would be rejected as impermissible by Zuraw while accepted by others.[24] And McCambridge et al. provide details of their study on alcohol consumption with the express purpose of starting a debate on the acceptability of the deception involved. We should not be surprised by the level of uncertainty regarding these studies, or by the level of disagreement that exists. The proposed distinction allows us to diagnose the cause of this uncertainty and disagreement. These are all cases where it is difficult to determine whether or not the deception involved is counterfactual-defeating or counterfactual-compatible. It is difficult to predict

[24] The original authors of the situationist example described above raise no concerns about the ethical status of their study. See Isen, Levin, *op. cit*. note 6.

whether or not participants would have consented to take part in the Milgram experiments or in the study on alcohol consumption had they been aware of the true aims of those studies. And it is difficult to predict whether or not subjects would have consented to participate in the situationist experiments had they been aware of what was taking place. This difficulty speaks in favour of the proposed distinction. Cases where there is uncertainty about the level of acceptability are also cases where it is difficult to determine whether or not the deception is counterfactual-compatible. This should give us some confidence that the counterfactual-defeating/counterfactual-compatible distinction is tracking some real and important feature of these cases. However, while this match speaks in favour of the accuracy of the proposed distinction, it is not good evidence of its practical usefulness. Cases where we most require guidance regarding the acceptability of a study are likely to be cases where we will struggle to determine what kind of deception is being carried out. How, then, can the proposed distinction actually further the debate on this issue?

The proposed distinction cannot provide a straightforward verdict in cases where we already struggle to determine the moral acceptability of a study. In contrast to rival suggestions, we cannot settle the matter simply by finding out whether the participants consented to the deception in advance, whether the deception focused on the aims or the existence of the study, or what the intentions of the researchers were when carrying out the deception. However, it is not correct to say that the proposed distinction is incapable of furthering the debate. This is because the distinction allows us to identify the *cause* of our uncertainty, and it suggests measures that can be taken to resolve it. Uncertainty about the acceptability of deception stems from uncertainty regarding whether the deception will be counterfactual-defeating. To resolve the uncertainty, therefore, we will require more evidence about whether the deception actually will be counterfactual-defeating. This evidence might be obtained in any number of different ways. For example, it may be possible to ask participants during debriefing whether or not they would have agreed to participate had they known the relevant facts. Milgram appears to have been sensitive to this way of defending the acceptability of his experiments. In later papers he finds it important to stress that 'Of the subjects, 83.7 per cent indicated that they were glad to have taken part in the study'.[25] Of course, receiving this information at the end of the study will not be ideal. The potentially problematic deception will already have taken place, and we may also worry about the judgement or candidness of those who have been recently deceived.[26] However, studying the feedback of those who have been subject to similar forms of deception may prove useful when contemplating future studies.

This will not be the only way of clearing up uncertainty regarding whether some potential deception will be counterfactual-defeating. For example, it would be possible for researchers to conduct a poll fully describing their planned experiment and asking whether or not people would be willing to take part. There is some evidence that McCambridge et al. are in favour of such an approach. In the process of justifying the deception involved in their alcohol consumption study they point out that 'student participants generally suggest that they do not mind taking part in deceptive studies.'[27] Again, any such evidence will be defeasible, and we ought to be cautious about this way of proceeding. Those surveyed would need to be provided with full information about the specific study, and we would need to be reasonably certain that they were representative of the likely participants in the actual study. Even then, any evidence will be less than definitive. It will always be possible that a given individual will be subject to counterfactual-defeating deception, even when all of the evidence suggests that the same deception would be counterfactual-compatible for most people. Perhaps the best that can be done is to ensure that certain ameliorating procedures are in place for whenever such a case is identified. One option would be to ensure that the individual is given the opportunity to withdraw their data from the study, and perhaps other possible options ought to also be considered. Either way, it is clear that procedures of this sort – asking for feedback from those who have participated in similar experiments and questioning large numbers of people about their willingness to take part in such a study – may allow researchers to become more certain that the deception involved in a given study will be counterfactual-compatible. If a level of justified certainty is reached then we ought to consider the deception used to be *prima facie* morally acceptable.

Of course, such a process may place additional burdens on researchers. For any study where the nature of deception is uncertain, researchers will be required to provide evidence for the claim that the deception will be counterfactual-compatible. This may require them to conduct surveys detailing the precise nature of the experiment and asking respondents whether or not they would be willing to take part in such a process. If the response indicates that consent would not be forthcoming then this

[25] S. Milgram. Some Conditions of Obedience and Disobedience to Authority. *Hum Relat* 1965; 18: 57–76, p. 58. Further evidence of this sort appears in S. Milgram. Issues in the Study of Obedience: A Reply to Baumrind. *Am Psychol* 1964; 19: 848–852. Questions regarding Milgram's use of debriefing are raised in Perry, *op. cit*. note 8, chapter 3.

[26] See D. Baumrind. Research Using Intentional Deception: Ethical Issues Revisited. *Am Psychol* 1985; 40: 165–174, p. 168–169.

[27] McCambridge et al., *op. cit*. note 3, p. 44.

will be one reason (stemming from considerations of deception) to disallow the study. It is important to be clear once again that worries about problematic deception are just one source of possible concern. Once we have evidence regarding whether the deception will be problematic, it is still possible that permission to proceed ought to be granted or withheld on some other grounds, such as levels of risk or lack of appropriate (actual) consent. The measures suggested here will only address the issue of whether considerations stemming from the use of deception render the study problematic. Regardless of the increased costs to the researcher in carrying out such measures, this process ought to be accepted as clearly preferable to either a system of absolute rejection of deceptive research (whereby many *prima facie* acceptable studies would be disallowed) or a system of absolute acceptance of deceptive research (whereby participants may be subjected to problematic deception). In this way, the distinction between counterfactual-defeating and counterfactual-compatible deception suggests a way forward in the current debate regarding the use of deception for the purposes of research.

CONCLUSION

I have attempted to show that the notion of counterfactual consent points towards a useful way of identifying problematic instances of deception. We ought to distinguish between those forms of deception that are counterfactual-defeating and those that are counterfactual-compatible. With this distinction in mind, it may be possible to advance the debate regarding the moral acceptability of various forms of deception for the purposes of research.

Acknowledgements

I would like to thank the anonymous reviewers for this journal for comments which helped me to improve this article, as well as members of the Bioethics Reading Group at the University of Edinburgh for helpful discussion on this general topic. I am grateful to the Arts and Humanities Research Council for generously funding the period of research during which this article was written.

Alan T. Wilson is a PhD candidate in philosophy at the University of Edinburgh. His current doctoral research focuses on virtue theory within ethics, with additional research interests including political philosophy and applied ethics.